Learning Heterogeneous Agent Collaboration in Decentralized Multi-Agent Systems via Intrinsic Motivation

Extended Abstract

Jahir Sadik Monon*
University of Dhaka
Dhaka, Bangladesh
jahirsadikmonon@gmail.com

Deeparghya Dutta Barua*
University of Dhaka
Dhaka, Bangladesh
deeparghya.csedu@gmail.com

Md Mosaddek Khan
University of Dhaka
Dhaka, Bangladesh
mosaddek@du.ac.bd

ABSTRACT

Multi-agent Reinforcement Learning (MARL) is emerging as a key framework for various sequential decision-making and control tasks. Unlike their single-agent counterparts, multi-agent systems necessitate successful cooperation among the agents. The real-world deployment of these systems requires decentralized training and execution (DTE), diverse agents, and learning from infrequent environmental rewards. These challenges become more pronounced under partial observability and the lack of prior knowledge about agent heterogeneity. While notable studies use intrinsic motivation (IM) to address reward sparsity or cooperation in decentralized execution settings, those dealing with heterogeneity typically assume centralized training for decentralized execution (CTDE). To overcome these limitations, we propose the CoHet algorithm, which utilizes a novel Graph Neural Network (GNN) based intrinsic motivation to facilitate the learning of heterogeneous agent policies in fully decentralized settings, under the challenges of partial observability and reward sparsity. Evaluation of CoHet in the Multi-agent Particle Environment (MPE) and Vectorized Multi-Agent Simulator (VMAS) benchmarks demonstrates superior performance compared to the state-of-the-art in a range of cooperative multi-agent scenarios. Our research is supplemented by an analysis of the impact of the agent dynamics model on the intrinsic motivation module, insights into the performance of different CoHet variants, and its robustness to an increasing number of heterogeneous agents.

KEYWORDS

Multi-agent Reinforcement Learning; Graph Neural Network; Intrinsic Rewards; Decentralized Training; Inter-agent Collaboration

ACM Reference Format:
Jahir Sadik Monon, Deeparghya Dutta Barua, and Md Mosaddek Khan. 2025. Learning Heterogeneous Agent Collaboration in Decentralized Multi-Agent Systems via Intrinsic Motivation: Extended Abstract. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 3 pages.

*Equal contribution.

Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

1 INTRODUCTION

The paradigm of Multi-agent Reinforcement Learning (MARL) is rapidly evolving to be pivotal in a broad spectrum of practical applications [5, 6, 8, 18]. Successfully executing tasks in many of these multi-agent scenarios requires the agents to learn collaboration while relying solely on local information and infrequent environmental rewards [19, 20]. The dependency on reward signals for the agents' learning process introduces the issue of reward sparsity [10] and the non-trivial nature of manually designing reward functions means that MARL systems need to be robust enough to deal with infrequent environmental rewards.

Applications such as package transport [9], disaster response [17], agriculture [12], etc. utilize agent heterogeneity such as distinct physical and behavioral traits of agents (e.g. speed, size, action space). Heterogeneity is also vital in multi-robot tasks as it enables efficient characterization and discovery of diverse behaviors, improving learning performance [14]. Moreover, as real-world applications constrain the agents to learn in a decentralized manner under partial observability, it is impractical for them to learn collaboration using a centralized algorithm with a global knowledge of the agents or the state space [11, 13]. Despite real-world requirements, existing solutions rely on global parameter sharing or centralized training (CTDE) [1] approaches. While Andres et al. [2], Zheng et al. [22] address heterogeneous agent collaboration under partial observability and reward sparsity, the former utilizes a central critic, and the latter defines heterogeneity differently as a mixture of on-policy, off-policy, and Evolutionary Algorithm agents.

In this work, we propose the **CoHet** algorithm to facilitate the learning of heterogeneous agent collaboration while addressing the constraints required for real-world applications, such as reward sparsity and partial observability. CoHet does not require any prior knowledge of agents' heterogeneity and is fully decentralized (DTE) [1]. Our specific contributions are as follows:

A Novel Intrinsic Reward Mechanism: To our knowledge, this is the first algorithm to calculate intrinsic rewards using a Graph Neural Network (GNN)-based local neighborhood observation aggregation. Unlike previous decentralized mechanisms that calculate intrinsic rewards based solely on single neighboring observations [16], this approach effectively captures neighborhood heterogeneity and provides more accurate reward estimations for diverse agent characteristics.

Integration with Established Algorithms: CoHet's intrinsic reward learning module can be integrated with existing decentralized policy learning algorithms with minimal adjustments, thus enhancing performance in cooperative MARL benchmarks.

We demonstrate this by incorporating the HetGPPO algorithm [4] which utilizes a GNN to aggregate observations for policy learning.

Extensive Validation and Robustness: We validate CoHet in the presence of heterogeneous agents in six different scenarios in the MPE [15] and VMAS [3] benchmarks, showing superior performance. In the full version of our paper, we also present findings on the impact of agent dynamics models on the intrinsic reward calculation, compare the two variants of the algorithm, and demonstrate its robustness to an increasing number of heterogeneous agents in a shared environment.

2 THE COHET ALGORITHM

CoHet is designed to improve decentralized cooperation among heterogeneous agents under sparse reward and partially observable scenarios. It integrates a GNN to aggregate local observations, ensuring geometric translation invariance by leveraging non-absolute features as node embeddings. Each agent collects only its locally accessible neighborhood observations, aggregates them, and utilizes GNN-learned node embeddings to calculate intrinsic rewards based on these aggregated local information. A forward dynamics model (Equation 2) is trained for each agent at regular intervals to predict the next observations based on current observations and actions. As shown in Equation 3, intrinsic rewards are computed by measuring the misalignment between an agent's predicted (computed using the forward dynamics model f_θ) and ground truth next aggregated neighborhood observations, with the misalignment penalized to encourage agents to refine their behavior to align better with their neighbors' predictions. These intrinsic rewards are weighted using Euclidean distances (Equation 1), prioritizing the influence of closer agents, and are combined with sparse extrinsic rewards to generate a dense reward signal for policy optimization.

$$w_j = \frac{(\|p_i - p_j\|)^{-1}}{\sum_{k \in \mathcal{N}_i^t \cap \mathcal{N}_i^{t+1}} (\|p_i - p_k\|)^{-1}} \tag{1}$$

$$\hat{o}_{j,i}^t = f_{\theta_j}(o_i^t, a_i^t) \tag{2}$$

$$r_{int_i}^t(o_i^t, a_i^t) = -\sum_{j \in \mathcal{N}_i^t \cap \mathcal{N}_i^{t+1}} w_j \times \|o_i^{t+1} - \hat{o}_{j,i}^t\| \tag{3}$$

CoHet introduces two key variants: $\text{CoHet}_{\text{team}}$ and $\text{CoHet}_{\text{self}}$. In $\text{CoHet}_{\text{team}}$, each agent shares its predictions of the next observations with its neighbors at the next time step, and agents learn to align their behaviors to minimize discrepancies between predicted and actual observations, fostering coordination at the team level. In contrast, $\text{CoHet}_{\text{self}}$ focuses solely on self-alignment, where each agent uses its own dynamics model to predict its future observations and optimizes behavior accordingly, leading to more independent decision-making at the cost of inter-agent collaboration. By leveraging local communication and prediction-based intrinsic rewards, CoHet enhances decentralized coordination without requiring centralized training, making it well-suited for complex multi-agent tasks with heterogeneity, sparse rewards, and partial observability. A more detailed and formal description of the CoHet algorithm is in the full version of our paper.

3 RESULTS

We demonstrate that both variants of CoHet ($\text{CoHet}_{\text{team}}$, $\text{CoHet}_{\text{self}}$) outperform the state-of-the-art decentralized heterogeneous MARL policy learning algorithm HetGPPO in each of the tasks evaluated on widely used VMAS and MPE benchmarks. The incorporation of the CoHet intrinsic reward module leads to the learning of collaborative behaviors among heterogeneous agents, evidenced by the improved performance over the baseline in these cooperative MARL tasks. We additionally compare CoHet with the SOTA MARL baseline, IPPO (Independent Proximal Policy Optimization), which is applicable in decentralized training settings for heterogeneous agents under partial observability similar to HetGPPO. Unlike the centralized critic-based heterogeneous policy learning approaches and widely used algorithms such as MADDPG [15], MAPPO [21], and COMA [7], these baselines along with CoHet address the more challenging problem of not relying on any centralized controller or prior knowledge of agent heterogeneity, but rather learning from only the locally observable information available to the diverse set of agents. As a result, to maintain uniform assumptions across methods, we show comparisons with the existing decentralized heterogeneous algorithms that operate under similar constraints. Furthermore, in our full paper, we analyze how each agent learns the dynamics model as time progresses and how it reduces the intrinsic reward penalty for misalignment. We compare the two variants of CoHet, analyze their performance, and demonstrate that the $\text{CoHet}_{\text{team}}$ variant is robust to an increasing number of heterogeneous agents in the shared environment, an issue encountered in previous methods [16].

Table 1: Mean Episodic Reward of CoHet variants vs. state-of-the-art baselines after 2×10^5 environment steps. Both CoHet variants simultaneously outperform the HetGPPO baseline in each task and outperform Independent PPO (IPPO) in four out of six tasks that require inter-agent cooperation

Scenario	IPPO	HetGPPO	$\text{CoHet}_{\text{team}}$	$\text{CoHet}_{\text{self}}$
Flocking	-0.73	-0.49	**0.41**	0.28
Navigation	**2.93**	0.75	1.97	1.80
Rev. Trans.	**7.92**	0.96	5.27	5.13
Sampling	26.13	17.81	**34.86**	31.75
Sim. Spread	-528.98	-701.15	-477.73	**-390.18**
Joint Pass.	-112.47	-55.10	**-2.73**	-9.11

4 DISCUSSION AND FUTURE WORK

We demonstrate that the GNN-based local neighborhood observation aggregation effectively models the neighborhood heterogeneity required for calculating prediction-based self-supervised intrinsic rewards. Future research can explore alternative intrinsic reward mechanisms (e.g. curiosity-driven, novelty-based) within decentralized heterogeneous MARL. Balancing intrinsic and extrinsic rewards remains an open challenge, and future work could investigate adaptive weighting mechanisms that prioritize agents with aligned sub-goals and heterogeneity types. Ultimately, we opine that future work on MARL cooperation should take the need for decentralized training and agent heterogeneity into account.

REFERENCES

[1] Christopher Amato. 2024. An Introduction to Centralized Training for Decentralized Execution in Cooperative Multi-Agent Reinforcement Learning. arXiv:2409.03052 [cs.LG] https://arxiv.org/abs/2409.03052

[2] Alain Andres, Esther Villar-Rodriguez, and Javier Del Ser. 2023. Collaborative training of heterogeneous reinforcement learning agents in environments with sparse rewards: what and when to share? *Neural Computing and Applications* 35, 23 (01 Aug 2023), 16753–16780. https://doi.org/10.1007/s00521-022-07774-5

[3] Matteo Bettini, Ryan Kortvelesy, Jan Blumenkamp, and Amanda Prorok. 2022. VMAS: A Vectorized Multi-Agent Simulator for Collective Robot Learning. In *International Symposium on Distributed Autonomous Robotic Systems*. Springer, 42–56.

[4] Matteo Bettini, Ajay Shankar, and Amanda Prorok. 2023. Heterogeneous Multi-Robot Reinforcement Learning. In *Proceedings of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS '23)*. International Foundation for Autonomous Agents and Multiagent Systems.

[5] Jeancarlo Arguello Calvo and Ivana Dusparic. 2018. Heterogeneous Multi-Agent Deep Reinforcement Learning for Traffic Lights Control.. In *AICS*. 2–13.

[6] Yongcan Cao, Wenwu Yu, Wei Ren, and Guanrong Chen. 2013. An Overview of Recent Progress in the Study of Distributed Multi-Agent Coordination. *IEEE Transactions on Industrial Informatics* 9, 1 (2013), 427–438. https://doi.org/10.1109/TII.2012.2219061

[7] Jakob Foerster, Gregory Farquhar, Triantafyllos Afouras, Nantas Nardelli, and Shimon Whiteson. 2018. Counterfactual Multi-Agent Policy Gradients. *Proceedings of the AAAI Conference on Artificial Intelligence* 32, 1 (Apr. 2018). https://doi.org/10.1609/aaai.v32i1.11794

[8] Taiki Fuji, Kiyoto Ito, Kohsei Matsumoto, and Kazuo Yano. 2018. Deep Multi-Agent Reinforcement Learning using DNN-Weight Evolution to Optimize Supply Chain Performance. https://doi.org/10.24251/HICSS.2018.157

[9] Brian Gerkey and Maja Mataric. 2002. Pusher-watcher: An approach to fault-tolerant tightly-coupled robot coordination. *Proceedings - IEEE International Conference on Robotics and Automation* 1, 464 – 469 vol.1. https://doi.org/10.1109/ROBOT.2002.1013403

[10] Joshua Hare. 2019. Dealing with sparse rewards in reinforcement learning. *arXiv preprint arXiv:1910.09281* (2019).

[11] Shariq Iqbal and Fei Sha. 2019. Actor-Attention-Critic for Multi-Agent Reinforcement Learning. In *Proceedings of the 36th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 97)*, Kamalika Chaudhuri and Ruslan Salakhutdinov (Eds.). PMLR, 2961–2970. https://proceedings.mlr.press/v97/iqbal19a.html

[12] Chanyoung Ju and Hyoung Son. 2019. Modeling and Control of Heterogeneous Agricultural Field Robots Based on Ramadge–Wonham Theory. *IEEE Robotics and Automation Letters* PP (09 2019), 1–1. https://doi.org/10.1109/LRA.2019.2941178

[13] Iou-Jen Liu, Raymond A Yeh, and Alexander G Schwing. 2020. PIC: permutation invariant critic for multi-agent deep reinforcement learning. In *Conference on Robot Learning*. PMLR, 590–602.

[14] Yuntao Liu, Yuan Li, Xinhai Xu, Yong Dou, and Donghong Liu. 2022. Heterogeneous Skill Learning for Multi-agent Tasks. In *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Eds.), Vol. 35. Curran Associates, Inc., 37011–37023.

[15] Ryan Lowe, Yi I Wu, Aviv Tamar, Jean Harb, OpenAI Pieter Abbeel, and Igor Mordatch. 2017. Multi-agent actor-critic for mixed cooperative-competitive environments. *Advances in neural information processing systems* 30 (2017).

[16] Zixian Ma, Rose Wang, Fei-Fei Li, Michael Bernstein, and Ranjay Krishna. 2022. Elign: Expectation alignment as a multi-agent intrinsic reward. *Advances in Neural Information Processing Systems* 35 (2022), 8304–8317.

[17] Nathan Michael, Shaojie Shen, Kartik Mohta, Vijay Kumar, Keiji Nagatani, Yoshito Okada, Seiga Kiribayashi, Kazuki Otake, Kazuya Yoshida, Kazunori Ohno, Eijiro Takeuchi, and Satoshi Tadokoro. 2012. Collaborative Mapping of an Earthquake Damaged Building via Ground and Aerial Robots, Vol. 92. https://doi.org/10.1007/978-3-642-40686-7_3

[18] Esmaeil Seraj, Rohan Paleja, Luis Pimentel, Kin Man Lee, Zheyuan Wang, Daniel Martin, Matthew Sklar, John Zhang, Zahi Kakish, and Matthew Gombolay. 2024. Heterogeneous policy networks for composite robot team communication and coordination. *IEEE Transactions on Robotics* (2024).

[19] Ceyer Wakilpoor, Patrick J. Martin, Carrie Rebhuhn, and Amanda Vu. 2020. Heterogeneous Multi-Agent Reinforcement Learning for Unknown Environment Mapping. arXiv:2010.02663 [cs.MA]

[20] Eric Wiewiora. 2010. *Reward Shaping*. Springer US, Boston, MA, 863–865. https://doi.org/10.1007/978-0-387-30164-8_731

[21] Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. 2022. The surprising effectiveness of ppo in cooperative multi-agent games. *Advances in Neural Information Processing Systems* 35 (2022), 24611–24624.

[22] Han Zheng, Pengfei Wei, Jing Jiang, Guodong Long, Qinghua Lu, and Chengqi Zhang. 2020. Cooperative heterogeneous deep reinforcement learning. *Advances in Neural Information Processing Systems* 33 (2020), 17455–17465.